ELSEVIER

Gene 275 (2001) 83–91

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Discovery and mapping of ten novel G protein-coupled receptor genes

Dennis K. Lee[a], Tuan Nguyen[b], Kevin R. Lynch[c], Regina Cheng[b], William B. Vanti[a], Oxana Arkhitko[a], Tressa Lewis[a], Jilly F. Evans[d], Susan R. George[a,b,e], Brian F. O'Dowd[a,b,*]

[a]*Department of Pharmacology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada*
[b]*The Centre for Addiction and Mental Health, Toronto, Ontario, M5S 2S1, Canada*
[c]*Department of Pharmacology, University of Virginia Health Sciences Center, Charlottesville, VA 22908, USA*
[d]*Department of Pharmacology, Merck & Co. Inc., 770 Sumneytown Pike, P.O. Box 4, West Point, PA 19486, USA*
[e]*Department of Medicine, University of Toronto, Toronto, Ontario, M5S 1A8, Canada*

Received 10 May 2001; accepted 1 August 2001
Received by W. Makalowski

Abstract

We report the identification, cloning and tissue distributions of ten novel human genes encoding G protein-coupled receptors (GPCRs) *GPR78*, *GPR80*, *GPR81*, *GPR82*, *GPR93*, *GPR94*, *GPR95*, *GPR101*, *GPR102*, *GPR103* and a pseudogene, *ψGPR79*. Each novel orphan GPCR (oGPCR) gene was discovered using customized searches of the GenBank high-throughput genomic sequences database with previously known GPCR-encoding sequences. The expressed genes can now be used in assays to determine endogenous and pharmacological ligands. *GPR78* shared highest identity with the oGPCR gene *GPR26* (56% identity in the transmembrane (TM) regions). *ψGPR79* shared highest sequence identity with the $P2Y_2$ gene and contained a frame-shift truncating the encoded receptor in TM5, demonstrating a pseudogene. *GPR80* shared highest identity with the $P2Y_1$ gene (45% in the TM regions), while *GPR81*, *GPR82* and *GPR93* shared TM identities with the oGPCR genes *HM74* (70%), *GPR17* (30%) and $P2Y_5$ (40%), respectively. Two other novel GPCR genes, *GPR94* and *GPR95*, encoded a subfamily with the genes encoding the UDP-glucose and $P2Y_{12}$ receptors (sharing >50% identities in the TM regions). *GPR101* demonstrated only distant identities with other GPCR genes and *GPR102* shared identities with *GPR57*, *GPR58* and *PNR* (35–42% in the TM regions). *GPR103* shared identities with the neuropeptide FF 2, neuropeptide Y2 and galanin GalR1 receptors (34–38% in the TM regions). Northern analyses revealed *GPR78* mRNA expression in the pituitary and placenta and *GPR81* expression in the pituitary. A search of the GenBank databases with the *GPR82* sequence retrieved an identical sequence in an expressed sequence tag (EST) partially encoding *GPR82* from human colonic tissue. The *GPR93* sequence retrieved an identical, human EST sequence from human primary tonsil B-cells and an EST partially encoding mouse *GPR93* from small intestinal tissue. *GPR94* was expressed in the frontal cortex, caudate putamen and thalamus of brain while *GPR95* was expressed in the human prostate and rat stomach and fetal tissues. *GPR101* revealed mRNA transcripts in caudate putamen and hypothalamus. *GPR103* mRNA signals were detected in the cortex, pituitary, thalamus, hypothalamus, basal forebrain, midbrain and pons.

Keywords: Orphan G protein-coupled receptor; Transmembrane; Intronless; Pseudogene; Chromosome

1. Introduction

As is frequently stated, GPCRs are the largest family of cell surface receptors and are responsible for the signal transduction for a diverse variety of ligands including nucleotides, biogenic amines, peptides and other small molecules (Marchese et al., 1999). GPCRs share a common heptahelical topography and these regions are embedded in the membrane. These seven transmembrane (TM) regions share the most significant levels of receptor identity. As a consequence, the majority of DNA sequences encoding GPCRs were found using methods dependent on sequence homology, mainly PCR or electronic sequence database screening (Marchese et al., 1998). GPCRs activated by similar ligands share the greatest identities with each other.

Abbreviations: aa, amino acid; BAC, bacterial artificial chromosome; EST, expressed sequence tag; GPCRs, G protein-coupled receptors; HTGS, high-throughput genomic sequences; nr, non-redundant; oGPCRs, orphan G protein-coupled receptors; ORF, open reading frame; TM, transmembrane

* Corresponding author. Department of Pharmacology, Medical Science Building, Room 4353, University of Toronto, 8 Taddle Creek Road, Toronto, Ontario, Canada M5S 1A8. Tel.: +1-416-978-7579; fax: +1-416-978-2733.
E-mail address: brian.odowd@utoronto.ca (B.F. O'Dowd).

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00651-5

However, newly discovered GPCRs frequently have only distant identities with known GPCRs, and these oGPCRs are difficult to characterize given the diversity in structure of the ligands and effector systems. This problem is compounded as we now realize that many endogenous ligands remain to be discovered. Increasingly, oGPCR characterization has utilized methods of 'reverse pharmacology', using the receptor as bait to retrieve ligands from tissue extracts. These efforts have identified endogenous ligands such as apelin, ghrelin, melanin-concentrating hormone, neuromedin U, the orexins, urotensin-II, and UDP-glucose (recently reviewed in Lee et al., 2001a; Civelli et al., 2001; Howard et al., 2001).

Approximately 250 mammalian genes encoding family A (or rhodopsin-like) GPCRs have been cloned (Lee et al., 2001a). As yet, the total number of cloned GPCRs reported in the literature including the secretin and metabotropic glutamate-like families of GPCRs falls short of the projected 616 GPCR-encoding sequences observed from the complete human genome sequence (Venter et al., 2001). Despite the human genome sequencing efforts, much work is still required to identify and clone the open reading frames (ORF) encoding the full complement of GPCR genes. Inserted into suitable expression vectors, these DNA sequences can be used to express the receptor in assays which will assist in ligand identification. For these reasons, we are continuing in our efforts to identify, catalog, compare and map the expression of GPCRs. We have recently reported the identification of the H4 histamine receptor (Nguyen et al., 2001), the cysteinyl leukotriene 2 receptor (Heise et al., 2000) and the oGPCR-encoding genes *GPR26*, *GPR57*, *GPR58* (Lee et al., 2000), *GPR61*, *GPR62*, *GPR63* and *GPR77* (Lee et al., 2001b). We now report the cloning of ten additional oGPCR-encoding genes named *GPR78*, *GPR80*, *GPR81*, *GPR82*, *GPR93*, *GPR94*, *GPR95*, *GPR101*, *GPR102* and *GPR103* as well as a pseudogene *ψGPR79*. *GPR78* and *GPR81* most closely resemble the oGPCR genes *GPR26* and *HM74*, respectively. *GPR80*, *GPR93* and *ψGPR79* shared highest identities with members of the purinoceptor family, while *GPR82* encoded an oGPCR distantly related to the purinoceptor-like oGPCR genes *GPR17* and *GPR34*. In addition, two novel genes *GPR93* and *GPR94* share significant identities with each other and with recently identified genes encoding the UDP-glucose (Chambers et al., 2000) and platelet ADP ($P2Y_{12}$) receptors (Hollopeter et al., 2001; Zhang et al., 2001), which together comprise a clustered family of genes on chromosome 3. *GPR101* shared distant identity with amine-binding GPCR genes, *GPR102* shared identity with the *PNR/GPR57/GPR58* amine receptor-like subfamily of GPCR genes and *GPR103* shared identities with peptide-binding receptors, including the neuropeptide FF 2, neuropeptide Y2 and galanin GalR1 receptors. We have also detected mRNA transcripts in tissues for *GPR78*, *GPR81*, *GPR94*, *GPR95*, *GPR101* and *GPR103*.

2. Materials and methods

2.1. Database searching

We queried the expressed sequence tag (EST) and high-throughput genomic sequences (HTGS) databases maintained by the National Center for Biotechnology Information with the amino acid (aa) sequences of various GPCRs using the TBLAST algorithm (Altschul et al., 1997). Retrieved sequences having statistically significant scores were further examined. The conceptualized protein sequences encoded by these sequences were used to query the non-redundant (nr) database to determine whether these sequences encoded previously known GPCRs.

2.2. GPCR gene and cDNA cloning

GPR78 was originally obtained in two fragments in an HTGS sequence from human chromosome 4 (GenBank Accession number: AC007104) which encoded the start methionine to the third intracellular loop (IC3) and from the carboxyl region of TM6 to the stop codon. Based on this sequence, two DNA fragments encoding from TM1 to TM4 and from TM6 to the stop codon were amplified from human genomic DNA. PCR products were extracted with phenol and chloroform, precipitated with ethanol and electrophoresed on a low-melting point agarose gel. Products in the expected size range were ligated into the *Eco*RV site of pBluescript SK(−) (Stratagene, La Jolla, CA) or $pcDNA_3$ (Invitrogen, Carlsbad, CA) and sequenced. Both fragments were observed to be identical to the HTGS sequence, and were used as probes to screen a human genomic library as previously described (Marchese et al., 1994). Library screening retrieved two phage DNA which encoded from the start methionine to IC3 and from the carboxyl region of TM6 to the stop codon. Primers designed upon TM5 (P1: 5′-GCTTCGTGCTGCCGCTG-3′) and TM7 (P2: 5′-CGGAG-CAGAGAGTACGTG-3′) were used to PCR amplify Marathon ready human fetal cDNA (Clontech, Palo Alto, CA) to retrieve a fragment sharing 100% identity in regions of overlap with the HTGS and human genomic phage DNA sequences. To obtain the complete intronless ORF of this gene, three overlapping segments encoding *GPR78* were obtained by PCR. Fragment 1 (encoding from the start methionine to TM5) was amplified from human genomic DNA using two primers (P3: 5′-GCGCCATGGGCCCCG-GCGAGG-3′, P4: 5′-GGTGACGGTGTCCATGCGC-3′). Fragment 2 was amplified using primers P1 and P2 from human fetal cDNA (described above). Fragment 3 (encoding from the third extracellular loop and extending 3′ of the stop codon) was amplified from human genomic DNA using two primers (P5: 5′-CTGGCGGAGCTCGTGCCC-3′, P6: 5′-GGCCAGTGCCCTTTCCAC-3′). These DNA fragments were joined by two further rounds of PCR. Round one consisted of fragments 1 and 2 together undergoing 30 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min using

primers P2 and P3. A second round of PCR amplified an aliquot of the first round with fragment 3 using primers P3 and P6 at similar cycling conditions. The PCR products were subcloned into the *Eco*RV site of pcDNA$_3$ (Invitrogen) and sequenced.

GPR103 was originally obtained in two overlapping fragments from the EST database (encoding from TM4 to TM7, GenBank Accession number: AI307658) and the HTGS database (encoding from TM6 to the stop codon, GenBank Accession number: AC005961). Analysis of the EST clone obtained from the I.M.A.G.E. Consortium revealed that this fragment encoded the receptor from TM2 to TM7. Based on these sequences, two DNA fragments encoding from TM3 to TM7 and from TM7 to the stop codon were obtained by PCR from human hypothalamus cDNA (Clontech) and human genomic DNA, respectively. PCR products were purified and ligated into pBluescript as described above. Sequence analysis revealed both DNA fragments to be identical with their respective database sequences. The fragment encoding TM3 to TM7 was used as a probe to screen a human hypothalamus cDNA library (Clontech), which retrieved a phage encoding *GPR103* from the start methionine to TM3 sharing 100% identity in the overlapping region with the EST derived sequence. To obtain a complete intronless ORF of this gene, the three overlapping fragments were joined by PCR as described above. The PCR products were subcloned into the *Eco*RV site of pcDNA$_3$ and sequenced.

To obtain DNA encoding other GPCRs, human genomic DNA was amplified by PCR using the following oligonucleotides: *ψGPR79* (5′-TGGGGCAGAGGCTGATGCCATGC-3′, 5′-AGCTGGATGCTCACCCAACTTGTTC-3′), *GPR80* (5′-GATTCATATTGCCAAACTGAAC-3′, 5′-CATCCTGAACATCTTAGGATG-3′), *GPR81* (5′-CTAACGCTCAGATAAGCATCTGTG-3′, 5′-GTCACCACTCTATCTTCCTCAGTG-3′), *GPR82* (5′-AATTCTATTCTAGCTCCTGTG-3′, 5′-CTAATAAAGTCACATGAATGC-3′), *GPR93* (5′-TTTGGCACGATGTTAGCC-3′, 5′-GTTCAGAGGGCGGAATCC-3′), *GPR94* (5′-AAGCAATGAACACCACAGTGATGC-3′, 5′-ATTATCTACGGAAGTCTCATC-3′), *GPR95* (5′-AGTTGGGTCTGTAAGGGAACC-3′, 5′-TTTATTTACACTTTGTACATATCG-3′), *GPR101* (5′-CTGGCTGTTGCCATGACGTCC-3′, 5′-GCCTTAGAACTAACTTCAAGG-3′), *GPR102* (5′-CAAACAACAAACAGCAGAACC-3′, 5′-CTTAGTGCTTAAACTTATTC-3′) and *P2Y$_{12}$* (5′-AAATAACCATCCTCTCTTTTGTTC-3′, 5′-CGAGTTCTGAACACAAAGAGATTG-3′). PCR conditions were as follows: denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 1 min. PCR products were purified, ligated into pcDNA3 or pBluescript vectors and sequenced as described above.

2.3. Northern blot analyses

Human and rat mRNA were extracted from various tissues as described previously (Marchese et al., 1994). Briefly, total RNA was extracted by the method of Chomczynski and Sacchi (1987), and poly (A)$^+$ RNA was isolated using oligo(dT) cellulose spin columns (Pharmacia, Uppsala, Sweden). RNA was denatured and separated by electrophoresis on a 1% formaldehyde agarose gel, transferred onto nylon membrane and immobilized by UV irradiation. The blots were hybridized with human GPCR-encoding ^{32}P-labeled DNA fragments, washed with 2× SSPE (3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA) and 0.1% SDS at 50°C for 20 min, washed again with 0.1× SSPE and 0.1% SDS at 50°C for 2 h and exposed to X-ray film at −70°C in the presence of an intensifying screen. In addition, two human Multiple Tissue Northern (MTN™) blots (Clontech) were used in *GPR78* (Human MTN Blot) and *GPR95* (Human MTN Blot IV) expression analyses according to the manufacturer's instructions.

3. Results

3.1. Cloning of GPCR-encoding genes

A search of the HTGS database with the *GPR26* sequence retrieved a novel human genomic sequence encoding a GPCR localized to chromosome 4 (GenBank Accession number: AC007104). The sequence was encoded on two fragments, from the start methionine to the middle region of IC3 and from the carboxyl region of TM6 to the stop codon. Primers were designed to PCR amplify human genomic DNA which retrieved fragments encoding TM1 to TM4 and from TM6 to the stop codon. These fragments were used to screen a human genomic library and two phage DNA fragments were retrieved. We also used TM5- and TM7-specific primers to amplify a DNA fragment from human fetal cDNA. The cDNA PCR product revealed 100% identity in regions overlapping the fragments previously obtained from genomic DNA, confirming these fragments as segments of the same gene. To obtain the full-length ORF, the three overlapping fragments were joined by PCR, and this clone was named *GPR78* (Fig. 1). *GPR78* encoded a 363 aa protein which shared extensive identities in the TM regions with *GPR26* (56%) (Table 1).

A customized search of the HTGS database retrieved a human genomic sequence (GenBank Accession number: AC021773) apparently encoding a novel GPCR. However, the sequence (*ψGPR79*) contained a frame-shift in the ORF. Primers were designed and used to PCR amplify this region of DNA sequence, and the product sequenced to verify the frame-shift in the TM5-encoding region confirming a pseudogene (*ψGPR79*). *ψGPR79* was used to search the HTGS database, which retrieved a related GPCR-encoding human genomic sequence localized to chromosome 13 (GenBank Accession number: AC026756). The PCR product containing this gene revealed an ORF of 336 aa in length, which we named *GPR80*. A search of the nr database with the projected *ψGPR79* aa sequence revealed significant identi-

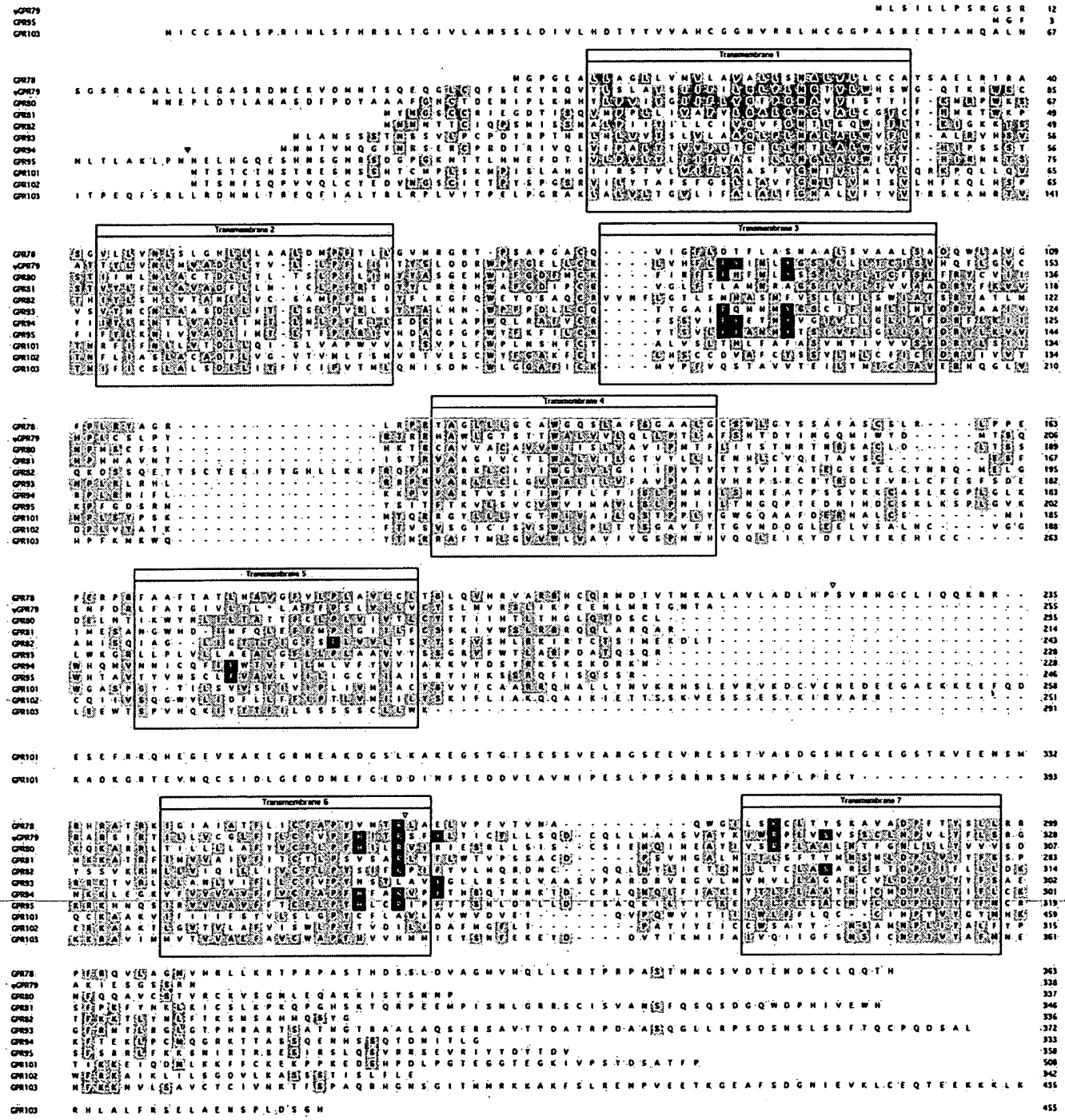

Fig. 1. oGPCR sequence alignments. Amino acid comparisons between novel GPCRs. The TM domain regions are indicated, and amino acids are numbered on the right. Black boxes with white lettering indicate conserved residues for purinergic binding (Erb et al., 1995; Jiang et al., 1997; Hoffmann et al., 1999). The presence of introns interrupting the *GPR78* and *GPR95* gene sequences is indicated by '▽' and '▼', respectively. The presence of a frame-shift interrupting the *ψGPR79* sequence is indicated by '*'. Residues shared between at least three aligned GPCR sequences are shaded.

ties in the TM regions with the purinoceptors $P2Y_2$ (51%), $P2Y_4$ (50%) and $P2Y_6$ (43%). A similar search with the *GPR80* sequence revealed greatest identities in the TM regions with the purinoceptors $P2Y_1$ (45%) and $P2Y_4$ (39%) and the cysteinyl leukotriene receptor CysLT2 (39%).

From the HTGS database we retrieved a GPCR-encoding sequence contained on a bacterial artificial chromosome (BAC) clone localized to chromosome 12q (GenBank Accession number: AC026331). The PCR amplified product (*GPR81*) encoded a 347 aa protein with identities in the TM regions with HM74 (70%), GPR31 (43%) and the purino-

Table 1
oGPCR sequence identities[a]

Receptor	% Identity	Accession number	Receptor	% Identity	Accession number
GPR78	GPR26, 49 (56) SSTR4, 21 (29) $5HT_6$, 20 (25)	AF411107	GPR95	UDP-glucose, 42 (62) $P2Y_{12}$, 39 (54) GPR94, 37 (51)	AF11114
GPR80	$P2Y_1$, 31 (45) $P2Y_4$, 29 (39) CysLT2, 31 (39)	AF411109	GPR101	RE2, <20 (31) $5HT_{1A}$, <20 (29) α_{1A}, <20 (29)	AF411115
GPR81	HM74, 50 (70) GPR31, 30 (43) $P2Y_1$, 23 (37)	AF411110	GPR102	PNR, 40 (42) GPR57, 36 (35) GPR58, 33 (35)	AF411116
GPR82	GPR34, 23 (31) GPR17, 24 (30) SSTR2, 23 (30)	AF411111	GPR103	NPFF2, 31 (38) NPY2, 29 (37) GalR1, 30 (35)	AF411117
GPR93	$P2Y_5$, 31 (40) GPR23, 28 (38) GPR17, 25 (36)	AF411112	ψGPR79	$P2Y_2$, 37 (51) $P2Y_4$, 37 (50) $P2Y_6$, 33 (43)	AF411108
GPR94	$P2Y_{12}$, 46 (57) UDP-glucose, 41 (52) GPR95, 37 (51)	AF411113			

[a] % Identity represents the shared sequence identities of each receptor with three of the closest GPCR sequences (TM identities are in parentheses). Each novel oGPCR sequence can be accessed through GenBank (http://www2.ncbi.nlm.nih.gov/genbank/query_form.html) with the listed Accession numbers.

ceptor P2Y1 (37%). The *HM74* sequence was retrieved in the same BAC clone as *GPR81*, indicating a clustering of these two genes.

GPR82 was retrieved using *GPR34* (from a search of the HTGS database) on a human BAC clone localized to chromosome 1 (GenBank Accession number: AL161458). The *GPR82* PCR product encoded a 336 aa protein sharing identities in the TM regions with the oGPCRs GPR34 (31%) and GPR17 (30%) and the somatostatin receptor SSTR2 (30%).

A search of the HTGS database using the cysteinyl leukotriene 2 receptor sequence retrieved a sequence (*GPR93*) encoding a GPCR on a human BAC clone localized to chromosome 12 (GenBank Accession number: AC006087). The PCR product obtained encoded a 372 aa protein which shared identities in the TM regions with the oGPCRs $P2Y_5$ (40%), GPR23 (38%) and GPR17 (36%).

A search of the HTGS database retrieved a cluster of three genes encoding GPCRs within the same contig localized to chromosome 3 (GenBank Accession number: AC024886). One of these genes was recently reported to encode the platelet ADP receptor, $P2Y_{12}$ (Zhang et al., 2001; Hollopeter et al., 2001). The $P2Y_{12}$ receptor gene shared homology with the other two genes, which we named *GPR94* and *GPR95*. A PCR product (*GPR94*) encoded a 333 aa protein which shared identities in the TM regions with the $P2Y_{12}$ receptor (57%), the UDP-glucose receptor (52%) and the receptor encoded by *GPR95* (51%). *GPR95* encoded a truncated GPCR, from TM1 to the stop codon, with an intron in the ORF evident by the lack of start methionine and the presence of an upstream in-frame stop codon. A search of the EST database retrieved one EST sequence from human testis mRNA encoding *GPR95* which also lacked a start methionine codon (GenBank Accession number: AA758208). This EST was acquired from the I.M.A.G.E. Consortium, sequenced, and confirmed to share 100% identity with the genomic sequence. A more recent database search retrieved an EST sequence from a human bladder cell line encoding *GPR95* with an alternative 5′ coding region (GenBank Accession number: BF028445), confirming a complete *GPR95* ORF and the presence of an intron between the start methionine and TM1-encoding regions. *GPR95* encoded a 358 aa protein which shared highest identities in the TM regions with the UDP-glucose receptor (62%), the $P2Y_{12}$ receptor (54%) and the receptor encoded by *GPR94* (51%).

We used the histamine H1 and H4 receptor sequences to retrieve two GPCR-encoding sequences, *GPR101* and *GPR102*, respectively, from the HTGS database localized to chromosome X (GenBank Accession number: AL390879) and chromosome 6 (GenBank Accession number: AL357505), respectively. A PCR amplified product (*GPR101*) encoded a 508 aa protein with identities in the TM regions with the oGPCR RE2 (31%), the serotonin $5HT_{1A}$ receptor (29%) and the α_{1A} adrenergic receptor (29%). The *GPR102* PCR product encoded a 342 aa protein sharing identities in the TM regions with the oGPCRs PNR (42%), GPR57 (35%) and GPR58 (35%).

A search of the EST and HTGS databases with the rat leukotriene LT2 receptor sequence retrieved human DNA sequences encoding a novel GPCR, GPR103. The novel receptor sequence was encoded on two overlapping fragments from a human kidney EST (encoding from TM2 to TM7) and from a human HTGS sequence (encoding from TM6 to the stop codon). Primers were designed to PCR amplify human hypothalamus cDNA and genomic DNA which retrieved two fragments encoding TM3 to TM7 and

from TM6 to the stop codon, respectively. The TM3 to TM7-encoding fragment was used to screen a human hypothalamus cDNA library, which retrieved a phage DNA fragment encoding GPR103 from the start methionine to TM3. The three DNA fragments revealed 100% identity in regions of overlap, confirming these fragments as segments of the same gene. To obtain the full-length ORF, the three overlapping fragments were joined by PCR, and this clone was named *GPR103*. *GPR103* encoded a 455 aa protein which shared identities in the TM regions with various peptide receptors, including the neuropeptide FF 2 (38%), neuropeptide Y2 (37%) and galanin GalR1 (35%) receptors.

3.2. Expression analyses

GPR78 mRNA transcripts were detected in human pituitary (1.1 kb) and placenta (two signals of 4.2 and 1.1 kb in size) (Fig. 2A,B). However, no *GPR78* transcripts were observed in human brain or specific CNS regions such as the frontal cortex, putamen, thalamus, hypothalamus, amygdala, hippocampus, pons, medulla and midbrain. In addition, human *GPR78* transcripts were also absent from skeletal muscle, lung, heart, liver, pancreas and kidney.

To determine the expression distribution of these novel oGPCRs, we performed Northern blots in various human and rat tissues. *ψGPR79* mRNA transcripts were not detected in human brain tissues, including frontal cortex, basal forebrain, pituitary, caudate nucleus, nucleus accumbens or hippocampus. In addition, Northern analyses did not reveal *ψGPR79* in rat brain, fetus, liver, spleen or adrenal gland tissue. Similarly, *GPR80* mRNA transcripts were not detected in human brain tissues including the frontal cortex, caudate putamen, thalamus, hypothalamus, hippocampus or pons tissue.

An mRNA transcript was detected for *GPR81* in human pituitary tissue, with an absence of signal in frontal, temporal and occipital lobes of the cortex, basal forebrain, caudate nucleus, nucleus accumbens, and hippocampus (Fig. 2C). Analyses of *GPR82* mRNA expression revealed no transcripts in human tissues, including regions of the CNS such as the frontal cortex, caudate putamen, thalamus, hypothalamus, hippocampus, pons and liver tissue. A search of the GenBank database retrieved an EST from cells derived from human colonic tissue encoding *GPR82* (GenBank Accession number: BF335802). The *GPR93* probe failed to detect mRNA transcripts in human frontal cortex, basal forebrain, caudate putamen, thalamus, or hippocampus. A search of the GenBank database revealed an EST encoding *GPR93* from human primary tonsil B-cells (GenBank Accession numbers: BF975186, BF663176 and BF129117) and an EST encoding a mouse *GPR93* orthologue expressed in the small intestine (GenBank Accession numbers: AV064817 and AV064680).

GPR94 mRNA transcripts of 3.2 kb were detected in human CNS tissues including the frontal cortex, caudate

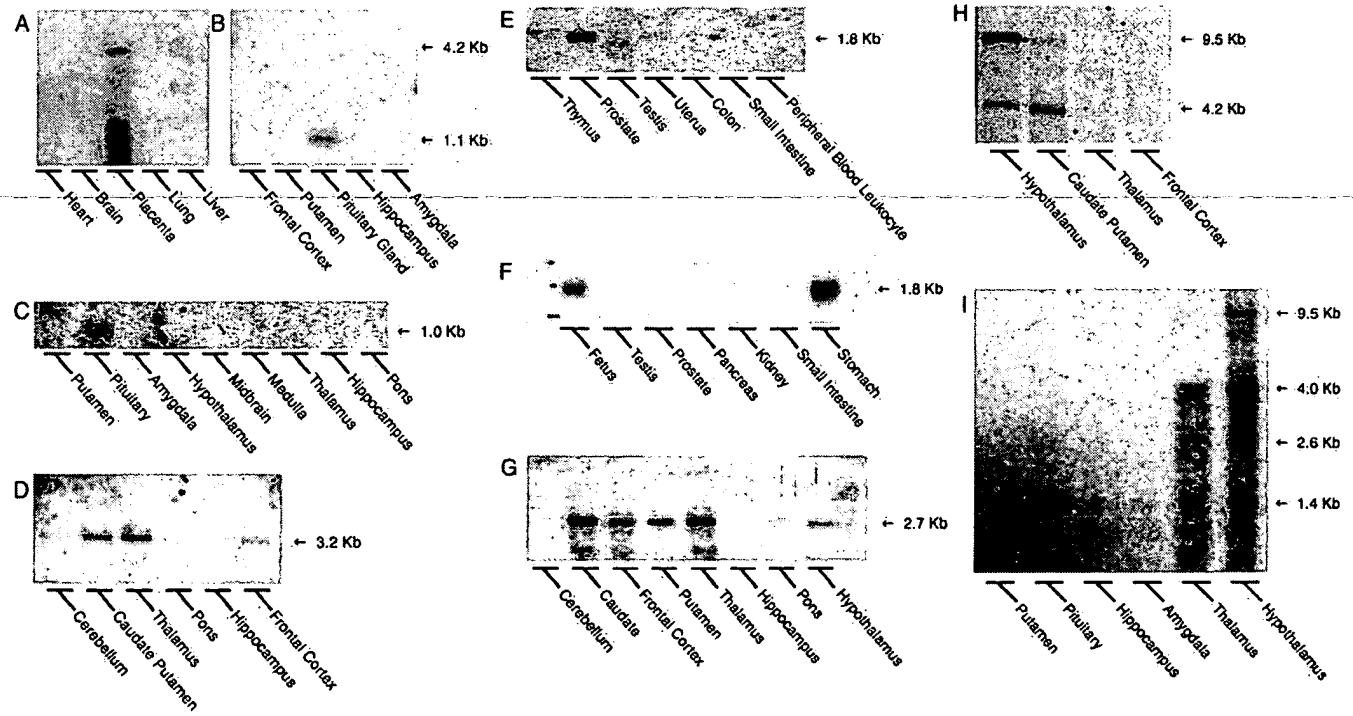

Fig. 2. oGPCR tissue distribution analyses. Northern blot of poly(A)$^+$ RNA (10 μg/lane), except for (A) and (E) which contained 2 μg/lane. (A) Human MTN™ and (B) human tissue distribution of *GPR78*. (C) *GPR81* and (D) *GPR94* human CNS tissue distribution. (E) Human MTN™ and (F) rat tissue distribution of *GPR95*, hybridized with a radiolabeled fragment encoding human *GPR95*. (G) *P2Y$_{12}$* and (H) *GPR101* human CNS tissue distribution. (I) Human tissue distribution of *GPR103*.

putamen and thalamus (Fig. 2D). *GPR94* was not detected in the hippocampus, pons or cerebellum and $P2Y_{12}$ was not detected in the hippocampus, cerebellum or in peripheral liver tissue. *GPR95* transcripts were observed in both human and rat peripheral tissue. A 1.8 kb signal was detected in human prostate, rat stomach and rat fetal tissues (Fig. 2E,F). However, *GPR95* expression was not observed in human thalamus, hypothalamus, hippocampus, pons or cerebellum or in rat whole brain tissue. In addition, we performed Northern analyses of *P2Y12* transcripts in human CNS tissues, which revealed faint signals of 2.7 kb in the frontal cortex, caudate putamen, thalamus, hypothalamus and pons (Fig. 2G). *GPR101* mRNA transcripts of 9.5 and 4.2 kb were detected in the caudate putamen and hypothalamus, with no expression detected in the frontal cortex, thalamus, hippocampus and pons (Fig. 2H). *GPR103* transcripts of 4.0, 2.6 and 1.4 kb were detected in the thalamus and hypothalamus, with a further 9.5 kb signal in the hypothalamus and a 1.4 kb signal in the pituitary (Fig. 2I). In addition, *GPR103* transcripts were also observed in the frontal and occipital cortices, basal forebrain, midbrain and pons (data not shown).

4. Discussion

Currently, ~350 human GPCRs have been cloned, as listed on the GPCRDB (G protein-coupled receptor database, http://www.gpcr.org/7tm/), with ~250 representing family A (or rhodopsin-like) GPCRs (Lee et al., 2001a). These receptors total approximately half the predicted 616 GPCR-encoding sequences contained in the human genome (Venter et al., 2001), although the veracity of this total number remains to be confirmed.

The identification of genes encoding the novel GPCRs predicts the existence of novel signaling systems leading to the discovery of novel ligands, as demonstrated by recent reports describing the discovery of apelin (Tatemoto et al., 1998), prolactin-releasing peptide (Hinuma et al., 1998), orexin (Sakurai et al., 1998), melanin-concentrating hormone (Bachner et al., 1999; Chambers et al., 1999; Lembo et al., 1999; Saito et al., 1999; Shimomura et al., 1999) and urotensin II (Ames et al., 1999; Liu et al., 1999; Mori et al., 1999; Nothacker et al., 1999) receptor-ligand systems. As these GPCR genes (many of which were cloned in our laboratory) were used in methods that led to the discovery and identification of these ligands (Lee et al., 2001a), we are continuing to isolate and characterize these novel genes.

We now report the discovery of ten novel oGPCRs and a pseudogene. *GPR78* is a paralogue of *GPR26*, apparent from shared identities (56% in the TM regions), a lack of asparagine-linked extracellular glycosylation sites, a short amino terminus, and similar gene structure (Fig. 1). *GPR78* and *GPR26* encoded receptors with shared conserved cationic arginine and lysine residues in TM6 and TM7, respectively, two residues recognized to play a role in purinergic binding and found only in P2Y receptors (Erb et al., 1995; Jiang et al., 1997) (Fig. 1). However, we reported calcium mobilization assays of human astrocytoma 1321N1 cells and *Xenopus laevis* oocytes transfected with *GPR26* were not responsive to nucleoside di- and tri-phosphates (Lee et al., 2000). We reported high levels of *GPR26* expression in many brain regions (Lee et al., 2000), while *GPR78* was detected only in the pituitary and placenta. However, the overall structural homology suggested that *GPR78* and *GPR26* may encode receptors that share a common endogenous ligand.

GPR80, *GPR81*, *GPR82*, *GPR93* and *ψGPR79* all shared identities to P2Y GPCRs or P2Y-like oGPCRs. Previously, at least three different nucleotide receptor phenotypes have been observed in mammalian tissue, including GPCRs activated by adenine nucleotides (e.g. $P2Y_1$ and $P2Y_{11}$), uridine nucleotides (e.g. $P2Y_6$) and by both adenine and uridine (e.g. $P2Y_2$ and $P2Y_4$) (King et al., 1998). *ψGPR79* shared closest identity with *P2Y2* and *P2Y4*, even though it does not encode a functional GPCR. *GPR80* was observed to share highest identity with $P2Y_1$, while *GPR81*, *GPR82* and *GPR93* shared identities with the P2Y-like oGPCR genes *HM74*, *GPR34* and *P2Y5*, respectively. Some of these oGPCR genes encode aa residues conserved amongst the P2Y receptors and shown to be involved in purinergic ligand binding (Erb et al., 1995; Hoffmann et al., 1999; Jiang et al., 1997) (Fig. 1). While expression was not observed for *GPR80*, *GPR82* and *GPR93* in various CNS regions, *GPR81* was observed to have an mRNA transcript in the pituitary, suggesting a role in neuroendocrine regulation.

The identities of *GPR94* and *GPR95* with the genes encoding the UDP-glucose and $P2Y_{12}$ receptors (>50% in the TM regions) indicate a novel subfamily of purinergic-like receptors. Previously, the UDP-glucose receptor was reported to have a distant sequence homology with the P2Y receptors, with an observed widespread tissue distribution in human brain and such peripheral tissues as placenta, adipose tissue, spleen, intestine, stomach, skeletal muscle, lung and heart (Chambers et al., 2000). The identification of the platelet ADP ($P2Y_{12}$) receptor resulted from cDNA isolated from rat platelets and human hypothalamus (Zhang et al., 2001; Hollopeter et al., 2001). An alignment of these receptors with the GPCRs encoded by *GPR94* and *GPR95* revealed several residues conserved in P2Y purinoceptors (see above) also conserved throughout this novel purinoceptor-like subfamily (Fig. 1). Northern analysis of *GPR94* mRNA revealed expression in various regions of the brain, suggesting a neuromodulatory role. In contrast, *GPR95* mRNA was detected in peripheral tissue (i.e. human prostate and rat stomach).

The receptor encoded by *GPR101* appeared to be a distant relative of the biogenic amine superfamily of GPCRs, with TM identities of ~30% with the adrenergic and serotonin receptors, as well as the muscarinic and dopamine receptors (data not shown). *GPR101* mRNA transcripts were observed in brain tissue, suggesting the presence of an endogenous

amine neurotransmitter ligand, perhaps novel in identity. The receptor encoded by *GPR102* shared significant TM identities with an amine binding receptor-like GPCR family (including PNR, GPR57 and GPR58) suggesting they may also share a common endogenous ligand. The receptor encoded by *GPR103* shared highest identities with several neuropeptide receptors. The significant levels of *GPR103* expression in the brain, particularly in the thalamus and hypothalamus, suggest an endogenous peptide ligand, perhaps involved in physiological functions such as pain modulation and neuroendocrine regulation.

Interestingly, some of these novel oGPCR genes appear to be clustered on various human chromosomes. The *GPR81* gene was localized to chromosome 12q, proximal to the closely related oGPCR gene, *HM74*. The *GPR82* gene was retrieved together with the *GPR34* gene in a human BAC clone localized on chromosome 1. Another search of the HTGS database revealed *GPR94*, *GPR95* and $P2Y_{12}$ localized within the same BAC clone on chromosome 3. This cluster of genes also includes the UDP-glucose receptor gene, which together with the more distant $P2Y_1$ receptor gene further localizes this cluster to chromosome 3q24-25 (interval D3S1279-1280) (Hollopeter et al., 2001). We have previously reported clusters of homologous GPCRs genes, including the *GPR40* through *GPR43* gene cluster (Sawzdargo et al., 1997) as well as the 5-HT_4-like pseudogene, *ψGPR57*, *GPR58* and *PNR* gene cluster (Lee et al., 2000). Given the significant sequence similarity of *GPR102* with *PNR*, *GPR57* and *GPR58* and its localization on chromosome 6, *GPR102* may be another paralogue member of this gene cluster.

In conclusion, we have identified ten novel GPCR genes and a pseudogene. Transcripts for *GPR78*, *GPR81*, *GPR94*, *GPR95*, *GPR101*, *GPR103* and $P2Y_{12}$ were detected in various CNS and peripheral tissues. Given the high levels of identity observed within paralogous oGPCR gene clusters, future efforts will likely discover common endogenous ligands for each of these novel GPCR subfamilies. The increasing number of oGPCRs that continue to be isolated with unique distribution profiles in brain and periphery is indicative that the search for novel transmitter ligands should be intensified. These efforts have the tremendous potential to uncover novel physiological roles for these as yet unknown receptor-transmitter signaling systems.

Acknowledgements

This study was funded by a Canadian Institutes of Health Research (CIHR) grant to S.R.G. and B.F.O. and by Merck Frosst Research Laboratories, Inc.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Ames, R.S., Sarau, H.M., Chambers, J.K., Willette, R.N., Aiyar, N.V., Romanic, A.M., Louden, C.S., Foley, J.J., Sauermelch, C.F., Coatney, R.W., Ao, Z., Disa, J., Holmes, S.D., Stadel, J.M., Martin, J.D., Liu, W.S., Glover, G.I., Wilson, S., McNulty, D.E., Ellis, C.E., et al., 1999. Human urotensin-II is a potent vasoconstrictor and agonist for the orphan receptor GPR14. Nature 401, 282–286.

Bachner, D., Kreienkamp, H., Weise, C., Buck, F., Richter, D., 1999. Identification of melanin concentrating hormone (MCH) as the natural ligand for the orphan somatostatin-like receptor 1 (SLC-1). FEBS Lett. 457, 522–524.

Chambers, J., Ames, R.S., Bergsma, D., Muir, A., Fitzgerald, L.R., Hervieu, G., Dytko, G.M., Foley, J.J., Martin, J., Liu, W.S., Park, J., Ellis, C., Ganguly, S., Konchar, S., Cluderay, J., Leslie, R., Wilson, S., Sarau, H.M., 1999. Melanin-concentrating hormone is the cognate ligand for the orphan G-protein-coupled receptor SLC-1. Nature 400, 261–265.

Chambers, J.K., Macdonald, L.E., Sarau, H.M., Ames, R.S., Freeman, K., Foley, J.J., Zhu, Y., McLaughlin, M.M., Murdock, P., McMillan, L., Trill, J., Swift, A., Aiyar, N., Taylor, P., Vawter, L., Naheed, S., Szekeres, P., Hervieu, G., Scott, C., Watson, J.M., et al., 2000. A G protein-coupled receptor for UDP-glucose. J. Biol. Chem. 275, 10767–10771.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Civelli, O., Nothacker, H., Saito, Y., Wang, Z., Lin, S.H., Reinscheid, R.K., 2001. Novel neurotransmitters as natural ligands of orphan G-protein-coupled receptors. Trends Neurosci. 24, 230–237.

Erb, L., Garrad, R., Wang, Y., Quinn, T., Turner, J.T., Weisman, G.A., 1995. Site-directed mutagenesis of P2U purinoceptors. Positively charged amino acids in transmembrane helices 6 and 7 affect agonist potency and specificity. J. Biol. Chem. 270, 4185–4188.

Heise, C.E., O'Dowd, B.F., Figueroa, D.J., Sawyer, N., Nguyen, T., Im, D.S., Stocco, R., Bellefeuille, J.N., Abramovitz, M., Cheng, R., Williams, J., Zeng, Z., Liu, Q., Ma, L., Clements, M.K., Coulombe, N., Liu, Y., Austin, C.P., George, S.R., O'Neill, G.P., et al., 2000. Characterization of the human cysteinyl leukotriene 2 receptor. J. Biol. Chem. 275, 30531–30536.

Hinuma, S., Habata, Y., Fujii, R., Kawamata, Y., Hosoya, M., Fukusumi, S., Kitada, C., Masuo, Y., Asano, T., Matsumoto, H., Sekiguchi, M., Kurokawa, T., Nishimura, O., Onda, H., Fujino, M., 1998. A prolactin-releasing peptide in the brain. Nature 393, 272–276.

Hoffmann, C., Moro, S., Nicholas, R.A., Harden, T.K., Jacobson, K.A., 1999. The role of amino acids in extracellular loops of the human P2Y1 receptor in surface expression and activation processes. J. Biol. Chem. 274, 14639–14647.

Hollopeter, G., Jantzen, H.-M., Vincent, D., Li, G., England, L., Ramakrishnan, V., Yang, R.-B., Nurden, P., Nurden, A., Julius, D.J., Conley, P.B., 2001. Identification of the platelet ADP receptor targeted by antithrombotic drugs. Nature 409, 202–207.

Howard, A.D., McAllister, G., Feighner, S.D., Liu, Q., Nargund, R.P., Van der Ploeg, L.H., Patchett, A.A., 2001. Orphan G-protein-coupled receptors and natural ligand discovery. Trends Pharmacol. Sci. 22, 132–140.

Jiang, Q., Guo, D., Lee, B.X., Van Rhee, A.M., Kim, Y.C., Nicholas, R.A., Schachter, J.B., Harden, T.K., Jacobson, K.A., 1997. A mutational analysis of residues essential for ligand recognition at the human P2Y1 receptor. Mol. Pharmacol. 52, 499–507.

King, B.F., Townsend-Nicholson, A., Burnstock, G., 1998. Metabotropic receptors for ATP and UTP: exploring the correspondence between native and recombinant nucleotide receptors. Trends Pharmacol. Sci. 19, 506–514.

Lee, D.K., Lynch, K.R., Nguyen, T., Im, D., Cheng, R., Saldivia, V.R., Liu, Y., Liu, I.S., Heng, H.H., Seeman, P., George, S.R., O'Dowd, B.F., Marchese, A., 2000. Cloning and characterization of additional

members of the G protein-coupled receptor family. Biochim. Biophys. Acta 1490, 311–323.

Lee, D.K., George, S.R., Evans, J.F., Lynch, K.R., O'Dowd, B.F., 2001a. Orphan G protein-coupled receptors in the CNS. Curr. Opin. Pharmacol. 1, 31–39.

Lee, D.K., George, S.R., Cheng, R., Nguyen, T., Liu, Y., Brown, M., Lynch, K.R., O'Dowd, B.F., 2001b. Identification of four novel human G protein-coupled receptors expressed in the brain. Brain Res. Mol. Brain Res. 86, 13–22.

Lembo, P.M., Grazzini, E., Cao, J., Hubatsch, D.A., Pelletier, M., Hoffert, C., St-Onge, S., Pou, C., Labrecque, J., Groblewski, T., O'Donnell, D., Payza, K., Ahmad, S., Walker, P., 1999. The receptor for the orexigenic peptide melanin-concentrating hormone is a G-protein-coupled receptor. Nat. Cell Biol. 1, 267–271.

Liu, Q., Pong, S.S., Zeng, Z., Zhang, Q., Howard, A.D., Williams Jr., D.L., Davidoff, M., Wang, R., Austin, C.P., McDonald, T.P., Bai, C., George, S.R., Evans, J.F., Caskey, C.T., 1999. Identification of urotensin II as the endogenous ligand for the orphan G-protein-coupled receptor GPR14. Biochem. Biophys. Res. Commun. 266, 174–178.

Marchese, A., Docherty, J.M., Nguyen, T., Heiber, M., Cheng, R., Heng, H.H., Tsui, L.C., Shi, X., George, S.R., O'Dowd, B.F., 1994. Cloning of human genes encoding novel G protein-coupled receptors. Genomics 23, 609–618.

Marchese, A., George, S.R., O'Dowd, B.F., 1998. Cloning of G protein-coupled receptor genes: the use of homology screening and the polymerase chain reaction. In: Lynch, K.R. (Ed.), Identification and Expression of G Protein-Coupled Receptors. Wiley-Liss, New York, pp. 1–26.

Marchese, A., George, S.R., Kolakowski Jr., L.F., Lynch, K.R., O'Dowd, B.F., 1999. Novel GPCRs and their endogenous ligands: expanding the boundaries of physiology and pharmacology. Trends Pharmacol. Sci. 20, 370–375.

Mori, M., Sugo, T., Abe, M., Shimomura, Y., Kurihara, M., Kitada, C., Kikuchi, K., Shintani, Y., Kurokawa, T., Onda, H., Nishimura, O., Fujino, M., 1999. Urotensin II is the endogenous ligand of a G-protein-coupled orphan receptor, SENR (GPR14). Biochem. Biophys. Res. Commun. 265, 123–129.

Nguyen, T., Shapiro, D.A., George, S.R., Setola, V., Lee, D.K., Cheng, R., Rauser, L., Lee, S.P., Lynch, K.R., Roth, B.L., O'Dowd, B.F., 2001. Discovery of a novel member of the histamine receptor family. Mol. Pharmacol. 59, 427–433.

Nothacker, H.P., Wang, Z., McNeill, A.M., Saito, Y., Merten, S., O'Dowd, B.F., Duckles, S.P., Civelli, O., 1999. Identification of the natural ligand of an orphan G-protein-coupled receptor involved in the regulation of vasoconstriction. Nat. Cell Biol. 1, 383–385.

Saito, Y., Nothacker, H.P., Wang, Z., Lin, S.H.S., Leslie, F., Civelli, O., 1999. Molecular characterization of the melanin-concentrating-hormone receptor. Nature 400, 265–269.

Sakurai, T., et al., 1998. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. Cell 92, 573–585.

Sawzdargo, M., George, S.R., Nguyen, T., Xu, S., Kolakowski, L.F., O'Dowd, B.F., 1997. A cluster of four novel human G protein-coupled receptor genes occurring in close proximity to CD22 gene on chromosome 19q13.1. Biochem. Biophys. Res. Commun. 239, 543–547.

Shimomura, Y., Mori, M., Sugo, T., Ishibashi, Y., Abe, M., Kurokawa, T., Onda, H., Nishimura, O., Sunino, Y., Fujino, M., 1999. Isolation and identification of melanin-concentrating hormone as the endogenous ligand of the SLC-1 receptor. Biochem. Biophys. Res. Commun. 261, 622–626.

Tatemoto, K., Hosoya, M., Habata, Y., Fujii, R., Kakegawa, T., Zou, M.X., Kawamata, Y., Fukusumi, S., Hinuma, S., Kitada, C., Kurokawa, T., Onda, H., Fujino, M., 1998. Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. Biochem. Biophys. Res. Commun. 251, 471–476.

Venter, J.C., et al., 2001. The sequence of the human genome. Science 291, 1304–1351.

Zhang, F.L., Luo, L., Gustafson, E., Lachowicz, J., Smith, M., Qiao, X., Liu, Y.H., Chen, G., Pramanik, B., Laz, T.M., Palmer, K., Bayne, M., Monsma, F., 2001. ADP is the cognate ligand for the orphan G-protein coupled receptor SP1999. J. Biol. Chem. 276, 8608–8615.